应用概率统计　第 33 卷
第 3 期　2017 年 6 月

Chinese Journal of Applied Probability and Statistics
Jun., 2017, Vol. 33, No. 3, pp. 310-316
doi: 10.3969/j.issn.1001-4268.2017.03.008

定数截尾数据缺失场合Frechet分布参数的近似极大似然估计 *

曾林蕊　　汤银才　　窦　雯

(华东师范大学统计学院, 上海, 200241)

摘　要: Frechet分布是一种重要的寿命分布, 本文给出了在定数截尾数据缺失场合两参数Frechet分布参数的近似极大似然估计, 并通过Monto-Carlo模拟说明了本文方法的可行性.

关键词: 两参数Frechet分布; 近似极大似然估计; 数据缺失

中图分类号: O212.1

英文引用格式: Zeng L R, Tang Y C, Dou W. Approximated maximum likelihood estimator for Frechet distribution under type II censoring [J]. Chinese J. Appl. Probab. Statist., 2017, 33(3): 310–316. (in Chinese)

§1. 引　　言

1927年, Frechet发表了第一篇关于极大值的渐近分布的论文[1]. Frechet分布在对自然灾害的分析上是非常有用的, 它是一种重要的寿命分布. 设产品寿命T服从两参数Frechet分布, 其分布函数和密度函数分别为

$$G(t,\alpha,\beta)=\exp\Big[-\Big(\frac{\beta}{t}\Big)^{\alpha}\Big],\quad g(t,\alpha,\beta)=\frac{\alpha}{\beta}\Big(\frac{\beta}{t}\Big)^{\alpha+1}\exp\Big[-\Big(\frac{\beta}{t}\Big)^{\alpha}\Big],\qquad t>0,$$

其中$\alpha>0$称为形状参数, $\beta>0$称为尺度参数.

关于Frechet分布的统计分析已有一些文献作了研究, Mann[2]讨论了Frechet分布和三参数威布尔分布的参数估计问题, Harlow[3]给出了Frechet分布在不同领域的应用, 指出它在工程应用中是一种重要的分布, Zaharim等[4]把Frechet分布应用到分析风速数据上, Abbas和Tang[5]研究了当形状参数已知时, Frechet分布中尺度参数的极大似然估计、贝叶斯估计和概率加权矩估计, Abbas和Tang[6]讨论了基于II型删失样本Frechet分布参数的极大似然估计和最小二乘估计.

在寿命试验中由于某种原因可能会发生部分数据缺失的情况, 关于数据缺失场合下的统计分析也有一些文献, 如田霆和刘次华[7]研究了定时截尾缺失数据下指数分布参数的近

*高等学校学科创新引智计划项目(批准号: B14019)和国家自然科学基金项目(批准号: 11271136、81530086)资助.
本文2016年5月10日收到, 2016年9月22日收到修改稿.

似极大似然估计, 王蓉华等[8]研究了在定数截尾数据缺失的场合下对数正态分布参数的近似极大似然估计, 徐晓岭[9]研究了数据缺失场合三参数威布尔分布的参数估计. 而本文则给出了在定数截尾数据缺失场合两参数Frechet分布参数的近似极大似然估计.

全文安排如下: 第二节研究了Frechet分布参数的近似极大似然估计, 由于参数的似然方程求解并非易事, 难以得到极大似然估计的显式解, 为克服上述不足, 本节导出了未知参数的近似似然方程, 给出了未知参数的显式估计量. 第三节进行了模拟计算并给出了结论.

§2.　Frechet分布参数的近似极大似然估计

令$X=\ln T$, 则X的分布函数和密度函数分别为

$$H(x)=\exp[-\mathrm{e}^{-(x-\mu)/\sigma}],\quad h(x)=\frac{1}{\sigma}\exp[-\mathrm{e}^{-(x-\mu)/\sigma}-(x-\mu)/\sigma],\qquad x\in R,$$

其中$\mu=\ln\beta$称为位置参数, $\sigma=1/\alpha$称为尺度参数.

再令$Z=(X-\mu)/\sigma$, 则Z的分布函数和密度函数分别为

$$F(z)=\exp[-\mathrm{e}^{-z}],\quad f(z)=\exp[-\mathrm{e}^{-z}-z],\qquad z\in R.$$

假定有n个产品进行寿命分布, 到有r个产品失效时停止试验, 其次序失效数据为

$$t_{(1)}\leqslant t_{(2)}\leqslant\cdots\leqslant t_{(r)},$$

若上述r个失效数据由于某种原因而使得有若干个数据缺失, 设剩下k个数据, 剩下的失效数据为

$$t_{(r_1)}\leqslant t_{(r_2)}\leqslant\cdots\leqslant t_{(r_k)},$$

令$z_i=(\ln t_i-\mu)/\sigma$, 则$z_{(1)}\leqslant z_{(2)}\leqslant\cdots\leqslant z_{(r)}$为来自分布函数和密度函数分别为$F(z)$、$f(z)$的前$r$个次序统计量, 剩下的$k$个失效数据为

$$z_{(r_1)}\leqslant z_{(r_2)}\leqslant\cdots\leqslant z_{(r_k)},$$

则似然函数为

$$\begin{aligned}L(\mu,\sigma)=c\sigma^{-k}[F(z_{(r_1)})]^{r_1-1}&\prod_{i=1}^{k-1}[F(z_{(r_{i+1})})-F(z_{(r_i)})]^{r_{i+1}-r_i-1}\\&\cdot[1-F(z_{(r_k)})]^{n-r_k}\prod_{i=1}^{k}f(z_{(r_i)}).\end{aligned}\tag{1}$$

对数似然函数为

$$\ln L(\mu,\sigma)=\ln c-k\ln\sigma+(r_1-1)\ln[F(z_{(r_1)})]+(n-r_k)\ln[1-F(z_{(r_k)})]$$

$$+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\ln[F(z_{(r_{i+1})})-F(z_{(r_i)})]+\sum_{i=1}^{k}\ln f(z_{(r_i)}).$$

令

$$\frac{\partial\ln(\mu,\sigma)}{\partial\mu}=0,\qquad \frac{\partial\ln(\mu,\sigma)}{\partial\sigma}=0,$$

于是得如下似然方程

$$\begin{aligned}&(r_1-1)\frac{f(z_{(r_1)})}{F(z_{(r_1)})}+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\frac{f(z_{(r_{i+1})})-f(z_{(r_i)})}{F(z_{(r_{i+1})})-F(z_{(r_i)})}\\&-(n-r_k)\frac{f(z_{(r_k)})}{1-F(z_{(r_k)})}-k+\sum_{i=1}^{k}\mathrm{e}^{-z_{(r_i)}}=0,\end{aligned}\tag{2}$$

$$\begin{aligned}&k+(r_1-1)\frac{f(z_{(r_1)})}{F(z_{(r_1)})}z_{(r_1)}+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\frac{f(z_{(r_{i+1})})z_{(r_{i+1})}-f(z_{(r_i)})z_{(r_i)}}{F(z_{(r_{i+1})})-F(z_{(r_i)})}\\&-(n-r_k)\frac{f(z_{(r_k)})}{1-F(z_{(r_k)})}z_{(r_k)}-\sum_{i=1}^{k}z_{(r_i)}+\sum_{i=1}^{k}z_{(r_i)}\mathrm{e}^{-z_{(r_i)}}=0.\end{aligned}\tag{3}$$

令

$$p_{r_i}=\frac{r_i}{n+1},\qquad q_{r_i}=1-p_{r_i},\qquad \xi_{r_i}=F^{-1}(p_{r_i}).$$

将函数$f(z_{(r_1)})/F(z_{(r_1)})$在点$\xi_{r_1}$处展开得

$$\frac{f(z_{(r_1)})}{F(z_{(r_1)})}\approx\alpha_1-\beta_1 z_{(r_1)},$$

其中

$$\alpha_1=\frac{f(\xi_{r_1})}{p_{r_1}}\Big[1+\xi_{r_1}(1-\mathrm{e}^{-\xi_{r_1}})+\frac{\xi_{r_1}f(\xi_{r_1})}{p_{r_1}}\Big],\qquad \beta_1=\frac{f(\xi_{r_1})}{p_{r_1}^2}[f(\xi_{r_1})+p_{r_1}(1-\mathrm{e}^{-\xi_{r_1}})].$$

将函数$f(z_{(r_k)})/[1-F(z_{(r_k)})]$在点$\xi_{r_k}$处展开得

$$\frac{f(z_{(r_k)})}{1-F(z_{(r_k)})}\approx\alpha_k+\beta_k z_{(r_k)},$$

其中

$$\alpha_k=\frac{f(\xi_{r_k})}{q_{r_k}}\Big[1+\xi_{r_k}(1-\mathrm{e}^{-\xi_{r_k}})-\frac{\xi_{r_k}f(\xi_{r_k})}{q_{r_k}}\Big],\qquad \beta_k=\frac{f(\xi_{r_k})}{q_{r_k}^2}[f(\xi_{r_k})-q_{r_k}(1-\mathrm{e}^{-\xi_{r_k}})].$$

将二元函数$[f(z_{(r_{i+1})})-f(z_{(r_i)})]/[F(z_{(r_{i+1})})-F(z_{(r_i)})]$在点$(\xi_{r_{i+1}},\xi_{r_i})$处泰勒展开得

$$\frac{f(z_{(r_{i+1})})-f(z_{(r_i)})}{F(z_{(r_{i+1})})-F(z_{(r_i)})}=\varepsilon_i+\omega_i z_{(r_i)}-\delta_i z_{(r_{i+1})},$$

其中

$$
\begin{aligned}
\varepsilon_i = & \frac{f(\xi_{r_{i+1}})-f(\xi_{r_i})}{p_{r_{i+1}}-p_{r_i}} \\
& + \xi_{r_{i+1}} f(\xi_{r_{i+1}}) \frac{(1-\mathrm{e}^{-\xi_{r_{i+1}}})(p_{r_{i+1}}-p_{r_i})+f(\xi_{r_{i+1}})-f(\xi_{r_i})}{(p_{r_{i+1}}-p_{r_i})^2} \\
& - \xi_{r_i} f(\xi_{r_i}) \frac{(1-\mathrm{e}^{-\xi_{r_i}})(p_{r_{i+1}}-p_{r_i})+f(\xi_{r_{i+1}})-f(\xi_{r_i})}{(p_{r_{i+1}}-p_{r_i})^2}, \\
\omega_i = & f(\xi_{r_i}) \frac{(1-\mathrm{e}^{-\xi_{r_i}})(p_{r_{i+1}}-p_{r_i})+f(\xi_{r_{i+1}})-f(\xi_{r_i})}{(p_{r_{i+1}}-p_{r_i})^2}, \\
\delta_i = & f(\xi_{r_{i+1}}) \frac{(1-\mathrm{e}^{-\xi_{r_{i+1}}})(p_{r_{i+1}}-p_{r_i})+f(\xi_{r_{i+1}})-f(\xi_{r_i})}{(p_{r_{i+1}}-p_{r_i})^2}.
\end{aligned}
$$

将二元函数$[z_{(r_{i+1})}f(z_{(r_{i+1})})-z_{(r_i)}f(z_{(r_i)})]/[F(z_{(r_{i+1})})-F(z_{(r_i)})]$在点$(\xi_{r_{i+1}},\xi_{r_i})$处泰勒展开得

$$
\frac{z_{(r_{i+1})}f(z_{(r_{i+1})})-z_{(r_i)}f(z_{(r_i)})}{F(z_{(r_{i+1})})-F(z_{(r_i)})} = a_i + b_i z_{(r_i)} - c_i z_{(r_{i+1})},
$$

其中

$$
\begin{aligned}
a_i = & \frac{\xi_{r_{i+1}} f(\xi_{r_{i+1}})-\xi_{r_i} f(\xi_{r_i})}{p_{r_{i+1}}-p_{r_i}} \\
& - \xi_{r_{i+1}} f(\xi_{r_{i+1}}) \frac{[1-\xi_{r_{i+1}}(1-\mathrm{e}^{-\xi_{r_{i+1}}})](p_{r_{i+1}}-p_{r_i})-[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2} \\
& - \xi_{r_i} f(\xi_{r_i}) \frac{[\xi_{r_i}(1-\mathrm{e}^{-\xi_{r_i}})-1](p_{r_{i+1}}-p_{r_i})+[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}, \\
b_i = & f(\xi_{r_i}) \frac{[\xi_{r_i}(1-\mathrm{e}^{-\xi_{r_i}})-1](p_{r_{i+1}}-p_{r_i})+[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}, \\
c_i = & -f(\xi_{r_{i+1}}) \frac{[1-\xi_{r_{i+1}}(1-\mathrm{e}^{-\xi_{r_{i+1}}})](p_{r_{i+1}}-p_{r_i})-[\xi_{r_{i+1}}f(\xi_{r_{i+1}})-\xi_{r_i}f(\xi_{r_i})]}{[p_{r_{i+1}}-p_{r_i}]^2}.
\end{aligned}
$$

将函数$\mathrm{e}^{-z_{(r_i)}}$在ξ_{r_i}处展开得

$$
\mathrm{e}^{-z_{(r_i)}} = d_i - f_i z_{(r_i)},
$$

其中

$$
d_i = \mathrm{e}^{-\xi_{r_i}} + \xi_{r_i}\mathrm{e}^{-\xi_{r_i}}, \qquad f_i = \mathrm{e}^{-\xi_{r_i}}.
$$

从而将(2)和(3)整理得到

$$
\begin{aligned}
& (r_1-1)(\alpha_1-\beta_1 z_{(r_1)}) + \sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(\varepsilon_i+\omega_i z_{(r_i)}-\delta_i z_{(r_{i+1})}) \\
& -(n-r_k)(\alpha_k+\beta_k z_{(r_k)}) - k + \sum_{i=1}^{k}(d_i - f_i z_{(r_i)}) = 0,
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
&k+(r_1-1)(\alpha_1-\beta_1 z_{(r_1)})z_{(r_1)}+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(a_i+b_i z_{(r_i)}-c_i z_{(r_{i+1})})\\
&-(n-r_k)(\alpha_k+\beta_k z_{(r_k)})z_{(r_k)}-\sum_{i=1}^{k} z_{(r_i)}+\sum_{i=1}^{k}(d_i-f_i z_{(r_i)})z_{(r_i)}=0.
\end{aligned}
\tag{5}
$$

化简(4)和(5)式求得参数σ、μ的AMLE

$$
\widehat{\sigma}=\frac{-D+\sqrt{D^2+4AE}}{2A},\qquad \widehat{\mu}=B-C\widehat{\sigma},
$$

从而可得α、β的AMLE: $\widehat{\alpha}=1/\widehat{\sigma}$, $\widehat{\beta}=\mathrm{e}^{\widehat{\mu}}$, 其中

$$
\begin{aligned}
M&=(r_1-1)\beta_1+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(\delta_i-\omega_i)+(n-r_k)\beta_k+\sum_{i=1}^{k} f_i,\\
MC&=(r_1-1)\alpha_1+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)\varepsilon_i-(n-r_k)\alpha_k-k+\sum_{i=1}^{k} d_i,\\
MB&=(r_1-1)\beta_1 x_{(r_1)}+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(\delta_i x_{(r_{i+1})}-\omega_i x_{(r_i)})\\
&\quad+(n-r_k)\beta_k x_{(r_k)}+\sum_{i=1}^{k} f_i x_{(r_i)},\\
A&=k+(r_1-1)(\alpha_1-\beta_1 C)C+C\sum_{i=1}^{k} d_i-C^2\sum_{i=1}^{k} f_i-kC\\
&\quad+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)(a_i+b_i C-c_i C)-(n-r_k)(\alpha_k+\beta_k C)C,\\
D&=(r_1-1)(\alpha_1-2\beta_1 C)(x_{(r_1)}-B)+\sum_{i=1}^{k}(d_i-1)(x_{(r_i)}-B)\\
&\quad-2C\sum_{i=1}^{k} f_i(x_{(r_i)}-B)-(n-r_k)(\alpha_k+2\beta_k C)(x_{(r_k)}-B)\\
&\quad+\sum_{i=1}^{k-1}(r_{i+1}-r_i-1)[b_i(x_{(r_i)}-B)-c_i(x_{(r_{i+1})}-B)],\\
E&=(r_1-1)\beta_1(x_{(r_1)}-B)^2+\sum_{i=1}^{k} f_i(x_{(r_i)}-B)^2+(n-r_k)\beta_k(x_{(r_k)}-B)^2.
\end{aligned}
$$

§3. 模　　拟

我们对$n=25$, 截尾数r分别为20和10, 参数真值取为$\alpha=1$, $\beta=1.5$时, 通过Monto-Carlo模拟的方法产生25个服从Frechet分布的随机数, 并进行了1 000次模拟, 利用本文方法可得Frechet分布参数的近似极大似然估计列于表1, 并把极大似然估计(用Laplace近似)列于表2, 用样本均值和均方误差(表1、表2中括号内的数字)考察参数估计的好坏. 在表1、表2中($r=20$, $k=5$)的时候, t_i表示非缺失的数据, 其它情况下的t_i表示缺失的数据.

表1 α、β的近似极大似然估计及均方误差

r	k	缺失情况	α	β
20	20	无缺失	1.0694 (0.0428)	1.5642 (0.1234)
	10	$t_2, t_4, t_6, t_8, t_{10}$ $t_{12}, t_{14}, t_{16}, t_{18}, t_{20}$	1.0840 (0.0520)	1.5650 (0.1332)
		t_1, t_4, t_5, t_8, t_9 $t_{11}, t_{12}, t_{15}, t_{18}, t_{19}$	1.0840 (0.0487)	1.5743 (0.1301)
	5	$t_1, t_4, t_8, t_{11}, t_{15}$	1.1133 (0.0766)	1.5597 (0.1374)
		$t_5, t_8, t_{13}, t_{16}, t_{20}$	1.1052 (0.0736)	1.5780 (0.1361)
10	10	无缺失	1.1784 (0.1571)	1.5006 (0.1528)
	8	t_3, t_6	1.1811 (0.1508)	1.5183 (0.1459)
		t_5, t_9	1.1994 (0.1660)	1.5088 (0.1551)
	5	$t_2, t_4, t_6, t_8, t_{10}$	1.2237 (0.1953)	1.5128 (0.1568)
		t_1, t_4, t_6, t_7, t_9	1.2088 (0.1942)	1.5154 (0.1755)

表2 α、β的极大似然估计(Laplace近似)及均方误差

r	k	缺失情况	α	β
20	20	无缺失	1.0660 (0.0422)	1.5407 (0.1164)
	10	$t_2, t_4, t_6, t_8, t_{10}$ $t_{12}, t_{14}, t_{16}, t_{18}, t_{20}$	1.0783 (0.0506)	1.5486 (0.1184)
		t_1, t_4, t_5, t_8, t_9 $t_{11}, t_{12}, t_{15}, t_{18}, t_{19}$	1.0737 (0.0471)	1.5495 (0.1114)
	5	$t_1, t_4, t_8, t_{11}, t_{15}$	1.1096 (0.0761)	1.5541 (0.1355)
		$t_5, t_8, t_{13}, t_{16}, t_{20}$	1.1063 (0.0710)	1.5604 (0.1375)
10	10	无缺失	1.1756 (0.1458)	1.5079 (0.1480)
	8	t_3, t_6	1.1800 (0.1779)	1.5120 (0.1472)
		t_5, t_9	1.1855 (0.1623)	1.5146 (0.1490)
	5	$t_2, t_4, t_6, t_8, t_{10}$	1.2276 (0.2300)	1.5125 (0.1493)
		t_1, t_4, t_6, t_7, t_9	1.2274 (0.2181)	1.5163 (0.1717)

结果表明:

1. 无论截尾数r是10还是20, 参数α和β的两种估计与真值比较接近, 均方误差也接近. 这表明近似极大估计与极大似然估计的精度是接近的, 但近似极大似然估计具有显然表示, 这是它的优势所在.

2. 当截尾数给定时, 数据缺失越多, 两种估计方法均表明α、β估计的偏差越大, 均方误差也越大, 说明估计的效果越差.

参 考 文 献

[1] Frechet M. Sur la loi de probabilité de l'écart maximum [J]. *Ann. Soc. Polon. Math.*, 1927, **6**, 93–116.

[2] Mann N R. Statistical estimation of parameters of the Weibull and Frechet distributions [M] // Tiago de Oliveira J. *NATO ASI Series, Vol. 131: Statistical Extremes and Applications*. Netherlands: Springer, 1984: 81–89.

[3] Harlow D G. Applications of the Frechet distribution function [J]. *Int. J. Mater. Prod. Tec.*, 2002, **17(5-6)**: 482–495.

[4] Zaharim A, Najid S K, Razali A M, et al. Analyzing Malaysian wind speed data using statistical distribution [C] // *Proceedings of the 4th IASME/WSEAS International Conference on Energy* & *Environment* (*EE'09*), *Cambridge, UK, February 24-26, 2009*. Stevens Point, Wisconsin, USA: WSEAS, 2009: 363–370.

[5] Abbas K, Tang Y C. Comparison of estimation methods for Frechet distribution with known shape [J]. *CJASR*, 2012, **1(10)**: 58–64.

[6] Abbas K, Tang Y C. Estimation of parameters for Frechet distribution based on type-II censored samples [J]. *CJASR*, 2013, **2(7)**: 36–43.

[7] 田霆, 刘次华. 定时截尾缺失数据下指数分布的参数AMLE [J]. 华侨大学学报(自然科学版), 2006, **27(4)**: 351–353.

[8] 王蓉华, 徐晓岭, 费鹤良. 对数正态分布参数的近似极大似然估计 [J]. 上海师范大学学报(自然科学版), 2000, **29(1)**: 39–43.

[9] 徐晓岭. 数据缺失场合三参数Weibull分布的参数估计 [J]. 上海师范大学学报(自然科学版), 2002, **31(2)**: 22–30.

Approximated Maximum Likelihood Estimator for Frechet Distribution under Type II Censoring

ZENG LinRui　TANG YinCai　DOU Wen

(*School of Statistics, East China Normal University, Shanghai 200241, China*)

Abstract: Frechet distribution is an important life distribution. In this paper, approximated maximum likelihood estimator for two parameter Frechet distribution under type II censoring is investigated. And the feasibility of this method is obtained through the Monto-Carlo simulation.

Keywords: two parameter Frechet distribution; approximated maximum likelihood estimator; type II censoring

2010 Mathematics Subject Classification: 62F10